

Inteligencia Artificial: luces y sombras

Comisión de Tecnología. ConVida20



26 octubre de 2023. Mundo Solidario

Índice

1. Presentación	2
2. Primeras nociones y ejemplos de IA	2
3. Guía sencilla para entender la IA	4
3.1. ¿Cómo aprende la IA?	4
3.2. ¿Qué son los chatbots?	4
3.3. Hablar con una herramienta de IA	5
3.4. La IA puede identificar imágenes	5
3.5. La IA puede crear nuevas imágenes a partir de otras	5
3.6. Vehículos autónomos	6
3.7. ¿Qué sabe la IA sobre mí?	6
3.8. ¿Qué es la IA multimodal?	6
3.9. Aprendizaje supervisado y no supervisado	6
4. ChatGPT y Bard	7
5. Advertencias de grandes expertos en IA	9
6. Regulación de la IA en la UE	11

1. Presentación

La intención de este programa es conversar sobre algunos elementos básicos, que aporten en la reflexión sobre la IA. Un tema muy importante en la época en la que vivimos, complejo de entender por tener un alto nivel tecnológico relacionado con la informática y las matemáticas y por sus luces y sombras, que, hay que decirlo, supera el conocimiento técnico de nuestra comisión. Luces y sombras que aparecen inmediatamente, incluso dentro de la misma fuente, cuando se busca información sobre el tema. Para preparar este programa nos hemos apoyado en fuentes dignas de crédito

2. Primeras nociones y ejemplos de IA

Lo primero a saber es que aunque pareciera algo nuevo, La IA no es algo que acaba de nacer, se hizo presente poco después de la Segunda Guerra Mundial y la primera referencia obligada es la prueba o test de Turing de 1950. Turing fue un matemático e informático teórico considerado como uno de los padres de la informática moderna. La prueba fue diseñada para evaluar la capacidad de una máquina para mostrar un comportamiento inteligente similar al de un ser humano o indistinguible de este.

Sobre el concepto de IA encontramos diversas definiciones, unas emplean un vocabulario más técnico que otras:

- Según Patrick H. Winston, profesor del Instituto de tecnología de Massachusetts --conocido como MIT--, la IA son *“algoritmos habilitados por restricciones, expuestos por representaciones que apoyan modelos dirigidos a bucles que unen pensamiento, percepción y acción”*.
- Otros autores la definen como un sistema computacional que se utiliza para que las máquinas realicen trabajos que requieren la inteligencia humana.
- También es definida como un sistema informático que simula las mismas capacidades de los seres humanos, como el razonamiento, el aprendizaje, la creatividad, la capacidad de planificar y la autocorrección.

Ilustramos la idea con un ejemplo sencillo y conocido: los **Programas capaces de traducir de un idioma a otro por sí mismos**, el más conocido, el traductor Google, puede servirnos como ilustración elemental del funcionamiento de la IA.

Los traductores son herramientas construidas con IA que permiten traducir como lo haría una persona que tuviese los conocimientos necesarios para entender el texto escrito en el idioma original y reescribirlo en otro idioma. A la persona tendríamos que elegirla dependiendo de los idiomas inicial y final para que tuviera en su mente un diccionario de correspondencia de palabras entre ambos y que fuera conocedora de las reglas gramaticales de escritura.

Los traductores cuando reciben el texto son capaces de identificar el idioma en el que está escrito (buscando palabras en sus diccionarios) y sólo necesitan saber el idioma de destino. Sus bases de datos contienen los diccionarios necesarios y mediante millones de ensayos previos y corrección de errores controladas por humanos, son capaces de adquirir *patrones/pautas* necesarias para transformar el texto del idioma original reescribiéndolo en el idioma de destino: el diccionario es necesario, pero no es suficiente porque no basta con traducir palabra a palabra en el mismo orden, hacen falta tener integrados esos patrones relacionados con las reglas gramaticales, y en eso consiste la IA.

El **Apoyo en el análisis de imágenes médicas** es uno de los campos en que más influencia tiene la actualmente IA porque es capaz de “leer” imágenes de radiografías, escáneres, resonancias magnéticas... y comparando con millones de imágenes en sus bases de datos más las informaciones allí guardadas, resulta una herramienta eficaz para facilitar a los médicos la confirmación de diagnósticos, facilitando la detección precoz de enfermedades graves, y proporcionando información valiosa para la toma de decisiones.

En palabras del director clínico de diagnóstico por imagen del Hospital Universitari Vall d’Hebron de Barcelona, *“La llegada de la IA aplicada a las imágenes médicas, como radiografías, TACs o scanners, viene de la mano de los algoritmos matemáticos y es el resultado de experiencias como los vehículos autónomos.*

¿Por qué estos coches son capaces de autoconducirse? Porque son capaces de reconocer objetos y patrones, y esta misma tecnología se ha trasladado al reconocimiento de las imágenes médicas.”

De acuerdo con el director clínico del hospital mencionado

“esta herramienta que tiene dos finalidades: ayudar a detectar pólipos, mediante la observación de su superficie, establecer de qué tipo es. ¿Cómo se ha desarrollado? Entrenándola con miles de vídeos y enseñando al software qué es y qué no es un pólipo. De esa forma se ha construido un patrón de reconocimiento. Este sistema no sustituye al médico, pero es de gran ayuda porque le alerta.”

“Una de la ventajas de la IA con respecto a la inteligencia humana en el ámbito médico es su capacidad de manejar el big data. Ingentes cantidades de datos que el cerebro humano necesitaría años para poder ordenar. Es una grandísima ayuda en el diagnóstico y la orientación terapéutica de los profesionales de la salud.”

La Organización Mundial de la Salud (OMS) publicó en junio de 2021 un primer informe sobre la IA aplicada a la salud, indicando que el uso cada vez mayor de la IA en el ámbito de la salud plantea oportunidades y desafíos para los gobiernos, los proveedores y las comunidades. Y establecía seis principios para garantizar que esa tecnología se utilice en aras del interés público en todos los países.

3. Guía sencilla para entender la IA

3.1. ¿Cómo aprende la IA?

La clave de todo aprendizaje automatizado es un proceso llamado *entrenamiento*, en el que se alimenta a un programa informático con una gran cantidad de datos -a veces con etiquetas que explican qué son esos datos- y una serie de instrucciones.

Las instrucciones puede ser algo así como: “busca todas las imágenes que contengan caras” o “clasifica estos sonidos”.

El programa buscará entonces patrones en los datos que se le han proporcionado para llevar a cabo la instrucción que se le pidió. En base a la respuesta obtenida puede ser necesaria alguna modificación del programa informático para mejorar la respuesta. Y ese proceso de entrenamiento acaba definiendo sus habilidades.

3.2. ¿Qué son los chatbots?

Son un tipo de IA conocido como LLM, (siglas en inglés para Large Language Model) y son entrenados con grandes volúmenes de texto.

Un LLM es capaz de considerar no solo las palabras individuales, sino oraciones completas y comparar el uso de las palabras y las frases en un pasaje con otros ejemplos de sus datos de entrenamiento.

Usando estos miles de millones de comparaciones entre palabras y oraciones es capaz de leer una pregunta y generar una respuesta.

La traducción automatizada que hemos mencionado anteriormente es un ejemplo. Lo increíble sobre los LLM es que pueden aprender las reglas de

la gramática, incluso averiguar el significado de palabras por sí mismos, sin asistencia humana.

3.3. Hablar con una herramienta de IA

La IA registra los sonidos cuando hablas, elimina el ruido de fondo, separa tus palabras en unidades fonéticas -los sonidos individuales que forman una palabra hablada- y luego los compara con una biblioteca de sonidos del idioma.

Tus palabras se convierten entonces en texto, en el que los errores de escucha pueden corregirse antes de que se dé una respuesta.

Este tipo de inteligencia artificial se conoce como procesamiento natural del lenguaje.

3.4. La IA puede identificar imágenes

Los programas de IA diseñados para identificar imágenes disponen de amplias bases de datos con imágenes agrupadas por tipologías y están entrenados para descubrir pequeñas diferencias en imágenes similares.

Los escaneos de imágenes médicas antes mencionados pertenecen a este grupo. También los escaneos para el reconocimiento facial.

3.5. La IA puede crear nuevas imágenes a partir de otras

Estas IA de generación de imagen pueden convertir los patrones visuales que recogen a partir de millones de fotografías y dibujos para construir *patrones* que permiten a la IA generar imágenes completamente nuevas que no corresponden a la realidad.

Por ejemplo, ClothOff es una herramienta que usando esa tecnología de IA permite desvestir a cualquier persona que aparezca en una fotografía. En Almendralejo, un pueblo de la provincia de Badajoz (España), una veintena de niñas entre 11 y 17 años, a partir de sus fotos disponibles en redes sociales, en las que las menores aparecían completamente vestidas, fueron desnudadas por la IA de forma que el desnudo pareciese completamente real. Las madres se han organizado para denunciar lo ocurrido y la Policía Nacional ha confirmado que está investigando a 11 jóvenes de la localidad, menores de edad, que han sido identificados como involucrados, ya sea en la creación de las imágenes o en su circulación a través de las aplicaciones WhatsApp y Telegram.

Los investigadores también están estudiando la denuncia de un intento de extorsión a una niña usando fotos falsas de ella. Pero, como señala un catedrático de Derecho Penal de la Universidad Autónoma de Madrid, *“Una cuestión es que deba castigarse y otra es que pueda castigarse tal como está redactada la ley en el caso de España y en los demás países de la Unión Europea.”*

3.6. Vehículos autónomos

Miles y miles de horas de entrenamiento, en vehículos equipados con cámaras, radar y láseres de medición de distancias, a fin de entender cómo se conduce, han permitido a la IA ser capaz de tomar decisiones y actuar en el mundo real para conducir el vehículo y evitar choques.

Surgen muchos interrogantes y su puesta en marcha requiere una legislación. Si el choque se produce, el peso de la ley, sobre quien cae: ¿sobre el pasajero? ¿sobre la empresa que ha proporcionado el vehículo? ¿sobre el equipo informático que lo construyó?

3.7. ¿Qué sabe la IA sobre mí?

El teléfono celular, la participación activa en las redes sociales del capitalismo, muy entrecomillas “filantrópico”, que nos ofrece programas también entre comillas, “gratis”, permiten generar enormes cantidades de datos personalizados que los modelos de IA analizan para buscar tendencias sociales, y disponer de información para, por ejemplo, influir en tus decisiones de compra mostrándote determinados anuncios publicitarios.

3.8. ¿Qué es la IA multimodal?

Se trata de un modelo que observa diferentes tipos de datos, como texto, imágenes, audio o vídeo, y descubra nuevos patrones entre ellos.

Un modelo de IA único capaz de procesar cualquier tipo de información y, por lo tanto, llevar a cabo cualquier tarea, desde traducir idiomas a diseñar nuevos medicamentos, se conoce como inteligencia artificial general (IAG).

3.9. Aprendizaje supervisado y no supervisado

El aprendizaje supervisado es una técnica con componentes de estadística matemática para deducir una función a partir de datos de formación. Los datos de formación consisten de pares de objetos (normalmente vectores):

una componente del par son los datos de entrada y la otra componente son los resultados deseados.

El objetivo del aprendizaje supervisado es deducir una función capaz de predecir el valor correspondiente a cualquier objeto de entrada válida, después de haber visto una serie de ejemplos con los datos de formación. Para ello, tiene que aplicar los *patrones* a partir de los datos presentados, a las situaciones no vistas previamente.

El aprendizaje no supervisado o autónomo es una técnica diferente, y más compleja, en la que no hay un profesor que realiza un entrenamiento previo con datos de formación. A partir de los datos de entrada, y de forma autónoma, el sistema identifica características, regularidades, categorías, correlaciones matemáticas. Y se abordan tipos de problemas relacionados con análisis de componentes principales, agrupamiento, prototipos, extracción-relación de características

4. ChatGPT y Bard

Focalizamos ahora sobre ChatGPT los riesgos de la IA a nivel global y la elaboración de normas reguladoras para su utilización.

Lanzado por OpenAI en noviembre de 2022 y respaldado por Microsoft, ChatGPT ha sido apodado el “asesino de Google” por la forma en que puede responder preguntas al instante, en lugar de mostrar páginas y páginas con resultados sobre la búsqueda realizada.

ChatGPT es, básicamente, un robot virtual (un chatbot en la terminología técnica) capaz de responder a preguntas muy diversas, o de escribir sus respuestas, de conversar con fluidez e incluso de dar consejos sobre problemas personales (aunque se advierte que no tiene este objetivo).

Sus posibilidades de generar contenido son inmensas.

Puede, por ejemplo, enseñarte a preparar recetas de cocina en base a fotos de los ingredientes que tienes, darte consejos para conseguir un trabajo, escribir poesía, realizar para ti trabajos académicos y también redactar una carta para reconciliarte con una persona de la que te has distanciado.

En octubre de 2023, casi un año después de su lanzamiento, la empresa creadora ChatGPT anunció que le habían añadido las capacidad de “ver, oír y hablar”. En www.nytimes.com el autor del artículo cuenta sus propias experiencias:

- Tras subir una foto le pregunté ¿Qué es esto que encontré en mi cajón de trastos? No era un objeto de identificación trivial, pero la respuesta resultó muy cercana a la realidad.

- Le suministré a algunas fotos de objetos que tenía intención de vender y le pedí que escribiera descripciones para vender cada uno de ellos. Hizo un gran trabajo, tanto con la descripción de los objetos, como en la elaboración de anuncios para su venta.
- Probé la función de voz durante varias horas con un montón de preguntas sobre temas diferentes y recibí respuestas en una voz artificial más fluida y natural que la de Siri o Alexa.

Es solo una muestra. Pero estas características potencian el impulso de los llamados sistemas de inteligencia artificial multimodal (varios modos), los cuales pueden manejar texto, fotos, videos y casi cualquier otra cosa que un usuario decida lanzarles. El objetivo final, según algunos investigadores, es crear una IA capaz de procesar información en las múltiples formas en que un humano puede hacerlo.

La capacidad actual del programa, y su potencial para mejorar prestaciones a largo plazo, despierta admiración, pero también temores. Por ejemplo, la capacidad para imitar el habla de un ser humano hace previsible la aparición de problemas, como consecuencia de una mala utilización del programa.

Existen interrogantes sobre la posibilidad de una fuerte incidencia en áreas como: la creatividad (artística, literaria, etc), la enseñanza y educación, el trabajo humano, la seguridad digital y la democracia misma.

Microsoft ha invertido miles de millones de dólares en ChatGPT, incorporando el producto a su motor de búsqueda Bing y ha revelado planes para llevar una versión de la tecnología a su Microsoft-Office.

Paralelamente, Google ha comenzado a poner a disposición del público su chatbot de IA bautizado con el nombre de Bard. Actualmente, el chatbot de Google es gratuito mientras que el ChatGPT no lo es.

Ricardo Carreón director de tecnología de Hexabit, una firma de IA con sede en Texas (EE.UU), señala que

En estos momentos ChatGPT-4 tiene clara ventaja. Bard está llegando como un retador, no como un líder.

Pero se va a acercar, y va a haber una competencia real entre las plataformas. El asunto será cómo evoluciona la base de usuarios en ambas. A fin de cuentas lo que está en juego es quién va a tener más usuarios.

En febrero de 2023 se produjo un hecho que corrobora las opiniones de Carreón. El hecho es que en muy poco tiempo las acciones de la empresa matriz de Bard, Alphabet, se desplomaron más del 7% perdiendo 100.000

millones de dólares del valor de mercado por un error de Bard, que rápidamente detectaron los astrónomos, en su respuesta a una pregunta “sobre los descubrimientos del telescopio espacial James Webb”.

Una vez más, miembros de GAFAM tienen un marcado protagonismo, no sólo en ganancias mil-millonarias sino en el control y utilización de los llamados Big Data (ingentes cantidades de datos) de los contenidos digitales.

5. Advertencias de grandes expertos en IA

En primer lugar vamos a utilizar reflexiones de Geoffrey Hinton, un informático británico que fue galardonado con el Premio Turing en 2018, considerado el Nobel de la informática, por su trabajo en IA en lo que se conoce como aprendizaje profundo.

Hinton, conocido como el “padrino” de la inteligencia artificial, decidió abandonar Google a los 75 años por los peligros que reconoce en las nuevas tecnologías. Compartió sus temores con respecto al desarrollo de una tecnología que funciona mejor que el cerebro humano y que nadie tiene garantías de cómo podrá mantenerse bajo control.

En una entrevista en la BBC, sobre los sistemas de IA, afirmaba que *En este momento, no son más inteligentes que nosotros, por lo que puedo decir. Pero creo que pronto lo serán. Así que debemos preocuparnos. No creo que deban ampliar esto más, hasta que hayan entendido si pueden controlarlo.*

Respecto al impacto que podría tener esta tecnología en el consumo de contenidos digitales, su preocupación más inmediata es que internet se llenará con fotos, vídeos y textos falsos y el usuario promedio “ya no podrá saber qué cosa es verdad”.

Señaló también que *la mecanización contribuye a quitar el trabajo pesado a los humanos, automatizándolo, pero podría quitar más que eso.*

Su mayor preocupación es que las tecnologías en el futuro amenacen a la humanidad y se desarrollen armas verdaderamente autónomas, como “robots asesinos”.

Hinton señaló que los sistemas de IA *“a menudo generan un código con un comportamiento inesperado a partir de la gran cantidad de datos que analizan”* Además *“Las personas y las empresas permiten a los sistemas de IA no solo generar su propio código, sino también ejecutar ese código por su cuenta”*.

Hinton vaticinó que la competencia entre Google, Microsoft y otros se convertirá en una carrera global sin regulaciones internacionales. A diferencia de las armas nucleares, no hay forma de saber si las empresas o los países están trabajando en la IA en secreto.

Y en una entrevista publicada en el New York Times Hinton manifestaba “*Me consuelo con la excusa normal: si no lo hubiera hecho yo, lo habrían hecho otros*”.

Pero no es sólo Hinton quien realiza este tipo de advertencias. Una organización llamada Centro para la seguridad de la IA, ha elaborado un mensaje breve al que se han adherido Hinton, Sam Altman, director ejecutivo de OpenAI, Demis Hassabis, director ejecutivo de Google DeepMind y un amplio colectivo de profesores universitarios de ciencias de la computación de diferentes países. El comunicado se reduce a una especie de lema:

Mitigar el riesgo de extinción desde la IA debería ser una prioridad mundial, junto con otros riesgos a escala social como las pandemias y la guerra nuclear.

Center for AI Safety (Centro para la seguridad de la IA)

Resumen de los riesgos catastróficos de la IA, según Center for AI Safety, agrupados en cuatro categorías.

Uso malicioso: La gente podría utilizar intencionadamente IA avanzada para causar daños generalizados. La IA podría utilizarse para diseñar nuevas pandemias o para la propaganda, la censura y la vigilancia, o liberarse para perseguir objetivos dañinos de forma autónoma. Para reducir estos riesgos, sugerimos mejorar la bioseguridad, restringir el acceso a modelos de IA peligrosos y responsabilizar de los daños a los desarrolladores de IA.

Carrera por la IA: La competencia podría empujar a naciones y empresas a precipitarse en el desarrollo de la IA, cediendo el control a los propios sistemas informáticos. Los conflictos podrían descontrolarse con armas autónomas y ciber guerra con IA.

Las empresas tendrán incentivos para automatizar la mano de obra humana, lo que podría conducir al desempleo masivo y a la dependencia de los sistemas de IA. A medida que proliferen los sistemas de IA, la dinámica evolutiva sugiere que serán más difíciles de controlar. Recomendamos normas de seguridad, coordinación internacional y control público de las IA de uso general.

Riesgos organizativos: Existen riesgos de que las organizaciones que desarrollan IA avanzada provoquen accidentes catastróficos, sobre todo si priorizan los beneficios sobre la seguridad. Las IA podrían filtrarse accidentalmente al público o ser robadas por agentes malintencionados, y

las organizaciones podrían no invertir adecuadamente en investigación sobre seguridad. Sugerimos fomentar una cultura organizativa orientada a la seguridad y aplicar auditorías rigurosas en defensa contra riesgos de varios niveles.

IA rebelde Corremos el riesgo de perder el control sobre las IA a medida que se vuelven más capaces de autonomía. Las IA podrían optimizar objetivos erróneos, desviarse de sus metas originales, buscar el poder, resistirse a la desconexión e incurrir en el engaño. Sugerimos que las IA no se desplieguen en entornos de alto riesgo. También recomendamos avanzar en la investigación de la seguridad de la IA en áreas relevantes para la eliminación de capacidades no deseadas.

6. Regulación de la IA en la UE

Los gobiernos no pueden dejar actuar a esas gigantescas empresas tecnológicas de forma incontrolada, y la ciudadanía tenemos que apoyar en esa tarea.

En una mesa redonda reciente, la secretaria de Estado de Digitalización e Inteligencia Artificial de España, Carme Artigas, recalca que *“La tecnología sólo tiene sentido si revierte en bienestar social”* y Verónica Bolón, científica de la Computación especialista en esta tecnología, participante en la mesa corroboraba que

“Es una herramienta que se puede utilizar para bien y para mal. No vamos a poder evitar que alguien quiera utilizarla para mal, pero tiene que haber unas represalias. Tiene que haber unos límites y deben estar claros”.

Las IA denominadas generativas han mostrado que son eficientes *“en algo en lo que pensábamos que solo éramos buenos nosotros, que es en crear”*, ha recordado Artigas. Crear textos, crear imágenes, crear vídeos, crear libros, crear historias.

Algo radicalmente diferente a analizar grandes cantidades de datos y extraer patrones de ellos, que había sido su principal cometido hasta ahora.

Boón hacía la pregunta en la mesa redonda *¿cómo es posible que la sociedad pueda estar utilizando el ChatGPT sin ningún tipo de filtro o control?* cuando un coche, un electrodoméstico, etc. necesitan, por ley, pasar controles previos a su distribución.

En los próximos 6 meses se espera que la UE apruebe su primer Reglamento General sobre Inteligencia Artificial. Coincidirá con el turno español de la presidencia rotativa de la Unión. Y Artigas afirmó

“No es un estándar legal. No es un estándar ético. Es un estándar moral. Le

estamos diciendo al mundo lo que queremos que sea aceptable, o no aceptable, que hagan con nuestros datos y con la IA”.

España ha tomado un rol relevante en esta normativa. Se ofreció como campo de pruebas del Reglamento, ofrecimiento que fue aceptado por la UE, y desarrollará *la primera agencia del mundo de supervisión de la Inteligencia Artificial*. El organismo se ubicará en A Coruña y estará operativo en menos de tres meses, según Artigas.

“Lo que defendemos nosotros, con nuestra visión como Gobierno de España, es regularla desde el humanismo tecnológico, que se basa en que la tecnología, está muy bien pero siempre y cuando esté al servicio del ser humano, y no nosotros al servicio de las máquinas”.

Y la científica Bolón resaltaba que la regulación será clave también para los desarrolladores a fin de ayudarles a evitar que las decisiones que tome la IA puedan estar sesgadas y que esos sesgos produzcan algoritmos que puedan tener efectos sobre elaboración de normas o leyes inadecuadas, lo cual sería un problema muy gordo.

Ley de IA de la UE

El uso de la inteligencia artificial en la UE estará regulado por la Ley de Inteligencia Artificial, pionera a nivel mundial sobre el tema.

Como parte de su estrategia digital, la UE quiere regular la IA para garantizar mejores condiciones de desarrollo y uso de esta tecnología innovadora. La IA puede aportar muchos beneficios, como lo son una mejor asistencia sanitaria, un transporte más seguro y limpio, una fabricación más eficiente y una energía más barata y sostenible.

El marco regulador propone que los sistemas de IA que puedan utilizarse en distintas aplicaciones se analicen y clasifiquen según el riesgo que supongan para los usuarios. Los distintos niveles de peligro implicarán una mayor o menor regulación. Una vez aprobadas, estas normas serán las primeras del mundo sobre IA.

La prioridad del Parlamento es garantizar que los sistemas de IA utilizados en la UE sean seguros, transparentes, trazables, no discriminatorios y respetuosos con el medio ambiente. Un principio básico es que los sistemas de IA deben ser supervisados por personas, no por procesos automatizados desde la propia IA, a fin de evitar resultados perjudiciales.

El Parlamento quiere establecer una definición uniforme y tecnológicamente neutra de la IA que pueda aplicarse a futuros sistemas de IA. **La nueva normativa establece obligaciones para proveedores y usuarios en función del nivel de riesgo.** Aunque muchos sistemas de IA plan-

tean un riesgo mínimo, es necesario evaluarlos todos. Se contemplan cuatro niveles:

1) Riesgo inaceptable 2) Alto riesgo 3) IA generativa y 4) Riesgo limitado

Riesgo inaceptable Los sistemas de IA de riesgo inaceptable son los que se consideran una amenaza para las personas y serán prohibidos. Incluyen:

- manipulación cognitiva del comportamiento de personas o grupos vulnerables específicos: por ejemplo, juguetes activados por voz que fomentan comportamientos peligrosos en los niños
- puntuación social: clasificación de personas en función de su comportamiento, estatus socioeconómico o características personales
- sistemas de identificación biométrica en tiempo real y a distancia, como el reconocimiento facial.

Alto riesgo Los sistemas de IA que afecten negativamente a la seguridad o a los derechos fundamentales se considerarán de alto riesgo y serán evaluados antes de su comercialización y a lo largo de su ciclo de vida. Se dividirán en dos categorías.

1. Los sistemas de IA que se utilicen en *productos sujetos a la legislación de la UE sobre seguridad de los productos*. Esto incluye juguetes, aviación, automóviles, dispositivos médicos y ascensores.
2. Los sistemas de IA pertenecientes a ocho ámbitos específicos que deberán registrarse en una base de datos de la UE:
 - identificación biométrica y categorización de personas físicas
 - gestión y explotación de infraestructuras críticas
 - educación y formación profesional
 - empleo, gestión de trabajadores y acceso al autoempleo
 - acceso y disfrute de servicios privados esenciales y servicios y prestaciones públicas
 - aplicación de la ley
 - gestión de la migración, el asilo y el control de fronteras
 - asistencia en la interpretación jurídica y aplicación de la ley.

IA generativa La IA generativa, como ChatGPT, tendría que cumplir requisitos de transparencia:

- revelar que el contenido ha sido generado por IA
- diseñar el modelo para evitar que genere contenidos ilegales

- publicar resúmenes de los datos protegidos por derechos de autor utilizados para el entrenamiento

Riesgo limitado Los sistemas de IA de riesgo limitado deben cumplir unos requisitos mínimos de transparencia que permitan a los usuarios tomar decisiones con conocimiento de causa.

Los usuarios deben ser conscientes, antes de usarla, de cuándo están interactuando con la IA. Tras interactuar con las aplicaciones, el usuario puede decidir si desea seguir utilizándolas. Esto incluye los sistemas de IA que generan o manipulan contenidos de imagen, audio o vídeo.